

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 18:15:11 ; Search time 18.37 Seconds
(Without alignments)
1199.035 Million cell updates/sec

Title: US-09-405-504A-53

Perfect score: 3384

Sequence: 1 MLLGASLVGVLLFSKLVKL.....RYVPLDQAYSRIQAGEEKL 643

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486.residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	2110	62.4	646	1	FATP_MOUSE	Q60714 mus musculus
2	2087	61.7	646	1	FATP_RAT	P97849 rattus norv
3	1064	31.4	620	1	VLCS_HUMAN	O14975 homo sapien
4	1044	30.9	620	1	VLCS_RAT	P97524 rattus norv
5	1022	30.2	620	1	VLCS_MOUSE	O35488 mus musculus
6	714	21.1	623	1	FAT1_YEAST	P38225 saccharomyc
7	350.5	10.4	522	1	CAIC_ECOLI	P31552 escherichia
8	318	9.4	661	1	ACSA_COPCI	O13440 coprinus ci
9	311.5	9.2	561	1	LCFA_ECOLI	P29212 escherichia
10	303	9.0	545	1	4CL1_SOLTU	P31684 solanum tub
11	300	8.9	545	1	4CL2_SOLTU	P31685 solanum tub
12	299	8.8	547	1	4CL1_TORAC	O24145 nicotiana t
13	293.5	8.7	561	1	4CL3_ARATH	Q98777 arabidopsis
14	286	8.5	544	1	4CL1_PETCR	P14912 petroselinu
15	282	8.3	544	1	4CL2_PETCR	P14913 petroselinu
16	278.5	8.2	553	1	4CL_VANPL	O24540 vanilla pla
17	271.5	8.0	562	1	LCFA_HAERIN	P46450 haemophilus
18	271	8.0	543	1	FAT2_YEAST	P38137 saccharomyc
19	270.5	8.0	656	1	ACSA_RHOCA	O68040 rhodobacter
20	267.5	7.9	542	1	4CL2_TORAC	O24146 nicotiana t
21	262	7.7	694	1	ACSA_CRIPIV	Q27156 cryptospori
22	260	7.7	5255	1	BACA_BACLI	O68006 b bacitraci
23	259	7.7	670	1	ACSA_EMENI	P16928 emericeila
24	259	7.7	3587	1	SRF1_BACSU	P27206 bacillus su
25	255.5	7.6	3587	1	SRF2_BACSU	Q04747 bacillus su
26	255	7.5	669	1	ACSA_PENCH	P36333 penicillium
27	253	7.5	6359	1	BACC_BACLI	O68008 b bacitraci
28	251	7.4	556	1	4CL2_ARATH	Q9S725 arabidopsis
29	250	7.4	3587	1	TYCB_BACBR	O30408 b tyrocidin
30	249	7.4	548	1	YDID_ECOLI	P38135 escherichia
31	247	7.3	675	1	ACSA_CANAL	O94049 candida alb
32	246	7.3	713	1	ACSL_YEAST	O1574 saccharomyc
33	246	7.3	2560	1	PPS2_BACSU	P39846 bacillus su

34	244	7.2	563	1	4CL1_ORYSA	P17814 oryza sativ
35	238.5	7.0	683	1	ACS2_YEAST	P52910 saccharomyc
36	238.5	7.0	1274	1	SRF3_BACSU	O08787 bacillus su
37	238	7.0	537	1	4CL_PINTA	P41636 pinus taeda
38	238	7.0	672	1	ACSA_PHYBL	Q01576 phycomyces
39	236.5	7.0	548	1	LUC1_LUCCR	O25686 helicobacte
40	236	7.0	662	1	ACSA_HELPY	P13129 luciola cru
41	235.5	7.0	6486	1	TYCC_BACBR	O30409 b tyrocidin
42	234	6.9	561	1	4CL1_ARATH	Q42524 arabidopsis
43	233	6.9	652	1	ACSA_ECOLI	P27550 escherichia
44	233	6.9	683	1	ACSL_KULI	O60011 kluyveromyc
45	231.5	6.8	660	1	ACSA_ALCEU	P31638 alcaligenes

ALIGNMENTS

RESULT 1
FATP_MOUSE
ID FATP_MOUSE STANDARD; PRT; 646 AA.
AC Q60714;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).
GN SLC27A1 OR FATP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95042740; PubMed=7954810;
RA Schaffer J.E., Lodish H.F.;
RT "Expression cloning and characterization of a novel adipocyte long
chain fatty acid transport protein.";
RL Cell 79:427-436(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98438516; PubMed=9765271;
RA Hui T.Y., Frohner B.I., Smith A.J., Schaffer J.E., Bernlohr D.A.;
RT "Characterization of the murine fatty acid transport protein gene and
its insulin response sequence.";
RL J. Biol. Chem. 273:27420-27429(1998).
CC -!- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS
ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
TRIGLYCERIDE SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SKELETAL MUSCLE, HEART
AND FAT. LOWER LEVELS IN BRAIN, KIDNEY, LUNG AND LIVER. NO
EXPRESSION IN SPLEEN OR INTESTINE.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement as long as its content is in no way
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: U15976; AAC71060.1;
DR EMBL: AF023258; AAC69640.1;
DR EMBL: AF023256; AAC69640.1; JOINED.
DR EMBL: AF023257; AAC69640.1; JOINED.
DR MGD: MGI:1347098; SLC27A1.
DR InterPro: IPR000873;
DR Pfam: PF0501; AMP-binding; 1.
DR PROSITE: PS00455; AMP_BINDING; 1.

KW Glycoprotein; Lipid transport; Transmembrane.

FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 140 160 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 537 557 POTENTIAL.
FT TRANSMEM 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 393 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 646 AA; 71276 MW; 910892BA8D985B4C CRC64;

Query Match 62.4%; Score 2110; DB 1; Length 646;

Best local similarity 60.9%; Pred. No. 8.9e-166;

Matches 391; Conservative 97; Mismatches 152; Indels 2; Gaps 2;

QY 4 GASLVGVLLFESKL-VLKLPWTQVGSLLFLYLGGSGWRIRVEIKTRDIRFGGLVLKV 62
DB 5 GAGTASVASLALLFWLGLPMTWSAAAFVYVGGGGFRLVCKTARDFGLSLVRV 64
QY 63 KAKVROCLQERRVPIIFASTVRRHDKTALIFEGTDTHTWTRQLDDEYSSVANFLQARG 122
DB 65 RLELRHRRAGDIPICFOAVRQPERLALVDASSGICWTFAQLDITYSNVANFLQGL 124
QY 123 LASGDVAALFPMENRNEFVGLWGLMAKGLVEALINTNLRDALLHCLTTSRARALVFGSE 182
DB 125 FAPGDVAVVLEGRPEFVGLWGLAKAGVVAALLNLRREPLAFCLGTSAKALIYGE 184
QY 183 MASACEVHASLDPSLSLFCSSGWEPAVPPSTEHLDPKADP-KHLPSCDPKGFDTKL 241
DB 185 MAAVAEVSQGLSKLKFCSDGLPESVLPDTQLDPLMLAEAPTTPLAQAPGKGMDDRL 244
QY 242 FYITSGTGLPKAAIIVVHSRYRMAALVYVYGFMRPNDIVDCLPLXHSAGNIYIGQC 301
DB 245 FYITSGTGLPKAAIIVVHSRYRMAALVYVYGFMRPNDIVDCLPLXHSAGNIYIGQC 304
QY 302 LLHGMVIRKKSASRFWDCKIKYNTIVQYIGELCYRLNQPPEAENQHOVRMALGN 361
DB 305 VIYGLTVVLRKKFSASRFWDCKYKNTVYQYIGELCYRLNQPPEAENQHOVRMALGN 364
QY 362 GLRQSIWTFSSRFHIQVAEFYCATCNCSLGNFDSQVGCNFSRLSVFYPIRLVRV 421
DB 365 GLRPAIMEEFTQRFVQICEFYGATECNCSIANMDKGVSGCFNSRLTHVPIRLVKV 424
QY 422 NEDTMELIRGDCVIPCQPGEGQLVGRILQKDPRLRRFDGYNQGANNAKAKDFKKG 481
DB 425 NEDTMEPLRDEGLICPCQPGEGQLVGRILQKDPRLRRFDGYNQGANNAKAKDFKKG 484
QY 482 DQAVLTGDLVMDGLVYFRDRTGDTFRWKGENVSTTEVEGTLRLLDMADVAVYGV 541
DB 485 DSAYLSDGLVMDGLVYFRDRTGDTFRWKGENVSTTEVEAVLSRLLGQTDVAVYGV 544
QY 542 PTEGRAGMAAVASPTGNCNCLERFAQVLEKELPLYARPIFLRLLPELHKTGYTKFQKTEL 601
DB 545 PGVEGRAGMAAIAIDPHSQDPSNMYELQKVLASVAPRIFLRLLPQVDTGTGTFKIQKTEL 604
QY 602 RKSGFDPAIVKDFLYLDAQKGYVPLDQEAYSRIQAGEKL 643
DB 605 QREGFDPRQTSRFLFLDLKQGRVYPLDERVHARICAGDFSL 646

RESULT 2

FATP_RAT

ID FATP_RAT STANDARD; PRT; 646 AA.

AC P97849;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).

GN SLC27A1 OR FATP.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98041635; PubMed=9375787;
RA Schaeap F.G., Hamers L., van der Vusse G.J., Glatz J.F.C.;
RL "Molecular cloning of fatty acid-transport protein cDNA from rat.";
RT Biochim. Biophys. Acta 1354:29-34(1997).
CC -!- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS
CC ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
CC AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
CC IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
CC TRIGLYCERIDE SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U89529; AAC53424.1; -.
DR InterPro; IPR000873; -.
DR Pfam; PF00501; AMP-binding; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Glycoprotein; Lipid transport; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 140 160 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 537 557 POTENTIAL.
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 393 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 646 AA; 71283 MW; 6450CF174CC2EB87 CRC64;

Query Match 61.7%; Score 2087; DB 1; Length 646;

Best local similarity 60.4%; Pred. No. 7.3e-166;

Matches 388; Conservative 98; Mismatches 154; Indels 2; Gaps 2;

QY 4 GASLVGVLLFESKL-VLKLPWTQVGSLLFLYLGGSGWRIRVEIKTRDIRFGGLVLKV 62
DB 5 GAGTASVASLALLFWLGLPMTWSAAAFVYVGGGGFRLVCKTARDFGLSLVRV 64
QY 63 KAKVROCLQERRVPIIFASTVRRHDKTALIFEGTDTHTWTRQLDDEYSSVANFLQARG 122
DB 65 RLELRHRRAGDIPICFOAVRQPERLALVDASSGICWTFAQLDITYSNVANFLQGL 124
QY 123 LASGDVAALFPMENRNEFVGLWGLMAKGLVEALINTNLRDALLHCLTTSRARALVFGSE 182
DB 125 FAPGDVAVVLEGRPEFVGLWGLAKAGVVAALLNLRREPLAFCLGTSAKALIYGE 184
QY 183 MASACEVHASLDPSLSLFCSSGWEPAVPPSTEHLDPKADP-KHLPSCDPKGFDTKL 241
DB 185 MAAVAEVSQGLSKLKFCSDGLPESVLPDTQLDPLMLAEAPTTPLAQAPGKGMDDRL 244
QY 242 FYITSGTGLPKAAIIVVHSRYRMAALVYVYGFMRPNDIVDCLPLXHSAGNIYIGQC 301
DB 245 FYITSGTGLPKAAIIVVHSRYRMAALVYVYGFMRPNDIVDCLPLXHSAGNIYIGQC 304
QY 302 LLHGMVIRKKSASRFWDCKIKYNTIVQYIGELCYRLNQPPEAENQHOVRMALGN 361
DB 305 VIYGLTVVLRKKFSASRFWDCKYKNTVYQYIGELCYRLNQPPEAENQHOVRMALGN 364
QY 362 GLRQSIWTFSSRFHIQVAEFYCATCNCSLGNFDSQVGCNFSRLSVFYPIRLVRV 421
DB 365 GLRPAIMEEFTQRFVQICEFYGATECNCSIANMDKGVSGCFNSRLTHVPIRLVKV 424


```

CC -----
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D85100; BA012722.1; -.
CC
CC DR INTERPRO; IPR000873; -.
CC DR Pfam; PF00501; AMP-binding; 1.
CC DR PROSITE; PS00455; AMP-BINDING; 1.
CC KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 107 127 POTENTIAL.
CC FT TRANSMEM 262 282 POTENTIAL.
CC SQ SEQUENCE 620 AA; 70693 MW; 6CF9362DC3805526 CRC64;

Query Match 30.9%; Score 1044; DB 1; Length 620;
Best Local Similarity 38.0%; Pred. No. 4.8e-79;
Matches 237; Conservative 113; Mismatches 241; Indels 32; Gaps 13;

QY 32 LYLGGGWRFIRVIKT-----IRRDFGGLVLLKAKVQCLOER--RTVPILFASTVR 85
DB 5 LYTGLAGLLPLLTCCCPYLLQDVRFYQLANMARVRSYRQRRPVRTILHVELEQAR 64

QY 86 RHPDKTALIFEGETDTHWTFQOLDEYSSSVANFLQAR-GLASGDVAAIFMENRNEFVGLWL 144
DB 65 KTHPKPFLFR--DETLTYAQVDRRSNOVARALHDHGLGRQDCVAFMGNEPAYVWLWL 122

QY 145 GMAKLGVEAALINTLRDALLHCLTTSRARALVFGSEMASAICEVHASL-DPSLSLFC 202
DB 123 GLKLGCPMACLNTNIRAKSLHCFQCCGAKVLLASPDQAEVEALPTLKKAIVVHS 182

QY 203 SGSEWEGAVPSTHELDPLK-APKHLSPCDKGTDFKLFYIYTSGETGLPKAAIVVHS 261
DB 183 SRTSNTNGVDTILDKVDGVSADPIPESWRS--EVTFTTTPAVYIYTSGETGLPKAAITNH 240

QY 262 RYRMAALVYGFPM-----RPNDIVYDCLPLYSAGNIVGIGQCLLHGMVWIRKFS 315
DB 241 R-----LWIGSLARSGIKAHDIYITMPLYHSAALMIGLHGCIVVGATFALRSKFS 293

QY 316 ASRFWDCCIKNCTIYQYIGELCRYLLNQPPEAENQHOVRMAGNGLRQSIWTFNSSRF 375
DB 294 ASQFWDCCRYNATVIQYIGELRLYLCNTPOKPNDRDHVKVIALGNLGRGVWREFIKRF 353

QY 376 HIPOVAEFYATGECNLSGDFSDQVAGCGFNSRLSFYPIRLVRVNETMELIRGPDGV 435
DB 354 GDHIVEFYASTEGNIGFMNYPKIGAVGRENLYQKKVVRHLLIKYDVEKDEPVRDANGY 413

QY 436 CIPQCGEPQGVRIQKDPLEFRFGYL--NOGANNKKIAKDVFKKQDAYLTGDVLMVD 494
DB 414 CIKPRGEVGLLICKITELTP---FPGYAGGKTQTEKKLRQVFKGDDVYFNSGDLMLD 470

QY 495 ELGYLPRDRGTDFRWKGNSTTEVEGTLRLLDMADVGVYGVPEGTGEGAGAAVA 554
DB 471 RENFIYFVRGDTFRWKGNVATTEVADIVGLVDFVEEVNRYGVDPVPGHEGRIGMASIK 530

QY 555 SPTG-NCDLERFAQVLEKELPLYARPIFLRLPELHKTGYKFOKTELKKEGDFPAIVKD 613
DB 531 MKNYFNGKGLFOHISEYLPYSRFRUKIQDTIETGTFRHKKVTLMEEGFNPSVIKD 590

QY 614 PLFYDAQKGRYVPLDQEAYSRI 636
DB 591 TLYFMDDETKTYVPMTEIDIINAI 613

RESULT 5
VCLS_MOUSE
ID VCLS_MOUSE STANDARD; PRT; 620 AA.

```

```

O35488;
AC 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
DE FATTY-ACID-COA LIGASE).
GN FACVL1 OR VLACS OR VLCS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; Tissue=Liver;
RA Kemp S., Lu J.-F., Smith K.D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF033031; AAB87982.1; -.
CC InterPro; IPR000873; -.
CC DR Pfam; PF00501; AMP-binding; 1.
CC DR PROSITE; PS00455; AMP-BINDING; 1.
CC KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 107 127 POTENTIAL.
CC FT TRANSMEM 262 282 POTENTIAL.
CC SQ SEQUENCE 620 AA; 70366 MW; 77C98BDD0DE3B9FFB CRC64;

Query Match 30.2%; Score 1022; DB 1; Length 620;
Best Local Similarity 37.7%; Pred. No. 3.2e-77;
Matches 235; Conservative 108; Mismatches 248; Indels 32; Gaps 12;

QY 32 LYLGGGWRFIRVIKT-----IRRDFGGLVLLKAKVQCLOER--RTVPILFASTVR 85
DB 5 LYTGLAGLLPLLTCCCPYLLQDVRFYQLANMARVRSYRQRRPVRTILHVELEQAR 64

QY 86 RHPDKTALIFEGETDTHWTFQOLDEYSSSVANFLQAR-GLASGDVAAIFMENRNEFVGLWL 144
DB 65 KTHPKPFLFR--DETLTYAQVDRRSNOVARALHDHGLGRQDCVAFMGNEPAYVWLWL 122

QY 145 GMAKLGVEAALINTLRDALLHCLTTSRARALVFGSEMASAICEVHASL-DPSLSLFC 202
DB 123 GLKLGCPMACLNTNIRAKSLHCFQCCGAKVLLASPDQAEVEALPTLKKAIVVHS 182

QY 203 SGSEWEGAVPSTHELDPL- LKADAPKHLSPCDKGTDFKLFYIYTSGETGLPKAAIVVHS 261
DB 183 SRTSNTNGVDTILDKVDGVSADPIPESWRS--EVTFTTTPAVYIYTSGETGLPKAAITNH 240

QY 262 RYRMAALVYGFPM-----RPNDIVYDCLPLYSAGNIVGIGQCLLHGMVWIRKFS 315
DB 241 R-----LWIGTGLAMSSGITAQDVITMPLYHSAALMIGLHGCIVVGATLALRSKFS 293

QY 316 ASRFWDCCIKNCTIYQYIGELCRYLLNQPPEAENQHOVRMAGNGLRQSIWTFNSSRF 375
DB 294 ASQFWDCCRYNATVIQYIGELRLYLCNTPOKPNDRDHVKVIALGNLGRGVWREFIKRF 353

QY 376 HIPOVAEFYATGECNLSGDFSDQVAGCGFNSRLSFYPIRLVRVNETMELIRGPDGV 435
DB 354 GDHIVEFYASTEGNIGFMNYPKIGAVGRENLYQKKVVRHLLIKYDVEKDEPVRDANGY 413

QY 436 CIPQCGEPQGVRIQKDPLEFRFGYL--NOGANNKKIAKDVFKKQDAYLTGDVLMVD 494
DB 414 CIKPRGEVGLLICKITELTP---FPGYAGGKTQTEKKLRQVFKGDDVYFNSGDLMLD 470

QY 495 ELGYLPRDRGTDFRWKGNSTTEVEGTLRLLDMADVGVYGVPEGTGEGAGAAVA 554
DB 471 RENFIYFVRGDTFRWKGNVATTEVADIVGLVDFVEEVNRYGVDPVPGHEGRIGMASIK 530

QY 555 SPTG-NCDLERFAQVLEKELPLYARPIFLRLPELHKTGYKFOKTELKKEGDFPAIVKD 613
DB 531 MKNYFNGKGLFOHISEYLPYSRFRUKIQDTIETGTFRHKKVTLMEEGFNPSVIKD 590

QY 614 PLFYDAQKGRYVPLDQEAYSRI 636
DB 591 TLYFMDDETKTYVPMTEIDIINAI 613

RESULT 5
VCLS_MOUSE
ID VCLS_MOUSE STANDARD; PRT; 620 AA.

```


RESULT	12
4CL1_TOBAC	
ID	4CL1_TOBAC STANDARD; PRT; 547 AA.
AD	O24145;
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA
DE	SYNTHASE 1).
DE	4CL1.
OS	Nicotiana tabacum (Common tobacco).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC	Solanales; Solanaceae; Nicotiana.
OX	NCBI_TaxID=4097;
RX	[1]
RP	SEQUENCE FROM N.A.
RY	MEDLINE=96416441; PubMed=8819324;
RA	Lee D., Douglas C.J.;
RT	"Two divergent members of a tobacco 4-coumarate:coenzyme A ligase
RT	(4CL) gene family, cDNA structure, gene inheritance and expression,
RL	and properties of recombinant proteins.";
RL	Plant Physiol. 112:193-205(1996).
CC	-!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC	4-COUMAROYL-COA.
CC	-!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC	METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC	-!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC - SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC

```
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U50845; AAB18637.1; -.
CC HSSP; P08659; 1LCI.
CC Mendel; 24477; Nicta;1179;24477.
CC InterPro; IPR000873; -binding; 1.
CC Pfam; PF00501; AMP-binding; 1.
CC PROSITE; PS00455; AMP_BINDING; 1.
CC Ligase; Phenylpropanoid metabolism; Multigene family.
KW SEQUENCE 547 AA; 59842 MW; 154DF6D684E3F51B CRC64;
SQ
Query Match      8.8%; Score 299; DB 1; Length 547;
Best Local Similarity 24.0%; Pred. No. 4.2e-17;
Matches 132; Conservative 93; Mismatches 232; Indels 92; Gaps
QY 90 KTALEFGTDTHTWTRQLDEYSVVANFLQARGLASGDVAALFMENRNEFWGLWGMAKL 149
Db | | | : | : | : | | | : | | | : | | | : | | | : | | | : |
46 RPCLINGANDQIVTYAEVELCTCRKAVGLNGLGIQQKDTINLLPNSPEVFVAFMGASYL 105
QY 150 GVEAALINTLRDALHCLLTTSRRALVTFGSEMASACEVHASLDPSLSFCSGSWEPG 209
Db | | | : | : | : | | | : | | | : | | | : | | | : | | | : |
106 GAISTMANPLFTPAEVGVKOAKASSAKIIITQSCFVGKVKYASEND--VKVICDSAPEG 163
QY 210 A-----VPPSTEHLDPALKDAPKHPSCPDKGFETDKLFYLYTSGTTGLPKAAIVVHSRY 264
Db | | | : | | | : | | | : | | | : | | | | | | | | | | | : |
164 CLHFSELATQSDEHEIPEVKIQPDVVALP-----YSSGTGLPKAGVMMLTHKGLV 212
QY 265 RMAALVYVG----FRMRPNDIVDCLPLYHSAGNIVIGQCLLHGM-----TVVIRKKFSA 316
Db | | | : | | | : | | | : | | | : | | | : | | | : | | | : |
213 TSAQAQVDGENGANLYMHSEDVLMCVLPFH----IYSLNSTILCGLRVGAAILIMQKFDI 268
QY 317 SRFWDCTKYNCITVOYIGELCYRLLNQPPEAENOHQVRMA-----LNGILRSQIWTN 370
Db | | | : | | | : | | | : | | | : | | | : | | | : | | | : |
269 AFLELIQKYKVSIGPFVPPIVLAIKSPIDVLSLVSFTVMVMSGAAPLGKELEDAVRTK 328
QY 371 FFSREHIPOVAEFYGATECNCSLG-----NFDSQVGACGNFSRIILSFVYPILRVVN 422
```

```

Query Match      8.8%; Score 299; DB 1; Length 547;
Best Local Similarity 24.0%; Pred. No. 4.2e-17;
Matches 132; Conservative 93; Mismatches 232; Indels 92; Gaps

QY 90 KTALEFGTDTHTTFRQLEDEYSSVANPLQARGLASGVAAIFMENNEFVGLWGLMAKL 149
   : || : : : : || : : : : || : : : : || : : : : || : : : : ||
Db 46 RPLCINGANDQIYTTAEVELTCRKVAVGLNKLGIQQQDTIMLLPNSPEVFVAFMGASYL 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 GVEAALINLRDRDALLHCLTTSRARALVFGSEMASALCEVHASLDPSLSLFCSGSWEPG 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 GAISTMANPLFTPAEIVKQAKASAKIIITQSCFVGKVOYASEND--VKVICDSAPEG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 A-----VPPSTEHLDPLLKDAKHLPCSPDKGFTDKLFYITVSGTGLPKAAIVVHSRY 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 CLHFSELTQSDHEIPEVKIQPDWVALP-----YSSGTTCLPGKVMLTHKGLV 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 RMAALVYVG----FMRRENDIVDCLPLYSAGNIVIGCOLHGM-----TVVIRKKFSA 316
   : | : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 TSVAQOVDGENANLYMHSESDVLMCVLPFH---IYSLNILLICGLRVGAAILIMQKFDI 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 SRFWDCTKYCNCTIVQYIGELCRLVLLNQPREARNQHVEMA-----LGNLGRQSIWTN 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 APFLELIQKYKVSIGPFPVPIVLAIAKSPIDVSLSVRTVMVSGAAPLGELEDAVRTK 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 FSSRHHIPOVAEYFGATECNSLGG-----NFDQVGACGNSRIILSFVPIRLVRVN 422

```

Db 329 FPN-----AKLGQCYGMEAGPVLAMCLAFAPKPFIDKSGACG-----TWVRNA 372
 QY 423 EDTMELIRGPDGVCIPQCPGEGQLVRIQKDLRRFDGYLNO--GANNKTKAKDVFKG 481
 Db 373 E--MKIVDPDTGSLP--RNQFGEI---CIRGDIQMK--GYLNDPEATRTIDK-----417
 QY 482 DQAYL--TGDVLMDELGYLYFDRDGTGTPRKWGENVSTTEVEGTLSRLDMDADVAVYGE 540
 Db 418 -EGWLHTGDIQFIDEDELFIYDRLKELIKYGFQVAPAEIEALLNHNPNISDAV---472
 QY 541 VPGTEGRAG---MAAVASPTGNCLE-REFAQVLEKELPLY---ARPIFLRLPELHKGT 593
 Db 473 VPMKDEQAGEVPAFVVRNGSNAITDEVKDFISKQVIFKVRKRVFFVETVP---KSPS 529
 QY 594 YKFOKTELR 602
 Db 530 GKILRKDLR 538

RESULT 13
 4CL3_ARATH STANDARD; PRT; 561 AA.
 AC 995777;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE 4-COUMARATE--COA LIGASE 3 (EC 6.2.1.12) (4CL 3) (4-COUMAROYL--COA
 DE SYNTHASE 3)
 GN 4CL3 OR ATIG65060 OR F16G16.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=99348176; PubMed=10417722;
 RA Ehling J., Buettner D., Wang Q., Douglas C.J., Somssich I.E.,
 RA Kombrink E.;
 RT "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana
 RL represent two evolutionarily divergent classes in angiosperms.";
 RL Plant J. 19:9-20(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu X., Liu Z.A., Luros J.S., Mafti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
 CC 4-COUMAROYL--COA.
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOIC
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to a license@isb-sib.ch).
 CC -----
 DR EMBL; AF106087; AD47194.1; -;
 DR EMBL; AF106088; AD47195.1; -;
 DR EMBL; AC009360; AAF06039.1; -;
 DR HSP; P08659; ILCI.
 DR InterPro; IPR000873; -;
 DR Pfam; PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Ligase; Phenylpropanoid metabolism; Multigene family.
 SQ SEQUENCE 561 AA; 61310 MW; C2EFF1C36B33F6DC CRC64;

Query Match 8.7%; Score 293.5; DB 1; Length 561;
 Best Local Similarity 25.3%; Pred. No. 1.3e-16;
 Matches 139; Conservative 84; Mismatches 237; Indels 89; Gaps 23;
 QY 89 DKTALIFEGDTHWTFQRLDEYSSVANFLQARGLASGDYAAIFMENNEFVGLWLGMAK 148
 Db 63 DKPCLIVGSTKSYTYGETHLICRRVASGLYKLGIRKGDVIMILLQNSAEVFSFGASM 122
 QY 149 LGVEAALINTNLRRDALLHCLTTSRARALVFGSEMAICEVHASLDPSLSLFCSGSWEP 208
 Db 123 IGAVSTANPYTSQELYKQLKSSGAKLIITHSQYVDKL---KNLGENLTITDEPTP 178
 QY 209 -GAVPPSTEHLDPILLKQAPKLPSCPKGTFDKLFYITTSGTGLPKAAIIVHSRYRMA 267
 Db 179 ENCLPFTSLITD---DETNPFTQVTDIGDDAAALPFSSTGTGLPKGVVLTHKSLITSV 234
 QY 268 ALVYVG---FRMRPNDIVDCLPLYSAGNIVGICQLLH---GMTVVRKFKFSARF 319
 Db 235 AQQVDGDNPNLYKSNVDILCVLPFH---IYSLNSVLLSLRSRGATVLLMHKFEIGAL 290
 QY 320 WDDCIKYNCTIVQYIGELCRYLLNPPREANQHOVRMA-----LGNLGRQSIWTFSS 373
 Db 291 LDLIQRHRTVAALVPLVIALAKNPTVNSYDLSSVFLSGAAPLKGELQDSL-----344
 QY 374 RHPHQ--VAEYFGATEC---NCSLG---NFDQVACGCFNSRILSFVYPIBLRVNE 423
 Db 345 RRRLPQAILGQYGMTEAGPVLMSLGFAPKPIPTKSGSCG-----TVVNAELKVVHL 398
 QY 424 DTMELIRG---PDGVCIPQCPGEGQLVRIQKDLRRFDGYLNO--GANNKTKAKDVFK 479
 Db 399 ET-RLSLGVNQPGEICR-----GQOIMKE-----YLNDEPATSATI-----434
 QY 480 KGDQAYL--TGDVLMDELGYLYFDRDGTGTPRKWGENVSTTEVEGTLSRLDMDADVAYG 538
 Db 435 -DEEGWLTGTGIVYDEDEDEIFIVDRLEKEVIFKFGQVPPAELESLLNHHISIAA--491
 QY 539 VEPVTEGRAG---MAAVASPTGNCLE-REFAQVLEKELPLYARPIFLRLPELHKGT 593
 Db 492 --VPONDEVAEGEVPVAFVVRNGN-DITEEDVKEVAKQVYFKLHKVFFVASIPKSPS 548
 QY 594 YKFOKTELR 602
 Db 549 GKILRKDLK 557

RESULT 14
 4CL3_PETCR STANDARD; PRT; 544 AA.
 AC P14912;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE 4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL--COA

DR Pfam; PF00501; AMP-binding; 1.
KW PROSITE; PS00455; AMP_BINDING; 1.
KN Ligase; Phenylpropanoid metabolism; Multigene family.
SQ SEQUENCE 544 AA; 59783 MW; B477965C68F8C534 CRC64;

Query Match 8.3%; Score 282; DB 1; Length 544;
Best Local Similarity 22.7%; Pred. No. 1.1e-15;
Matches 123; Conservative 98; Mismatches 244; Indels 78; Gaps 20;
QY 89 DKATLIEGTDTHWTFQDLDEYSSVANFLQAGLSDGDAIFAENRNEFFVGLWLGNAK 148
DB 41 DKSLINGATGETFTYSVELLSKRVASGLNKLQIQGGDTIMLLPNSPEYFFAFLGASY 100
QY 149 LGVEAALINLRDALLHCLTTSRRALVFGSEMAAICEVHA-----SLDPSLSLFCSG 204
DB 101 RGAISTWANFFTSAEVIQKASLAKLII-----TQACYVDKYDYAAEKNIQIICID 154
QY 205 SWEFGAVPSTEDHLDPLDKAPHLSPCKDGTGKLFYIYSGTGLPKAAIVVHSRY 264
DB 155 D-----APQDLHFSKLMEADESEMPVVIDS--DDVVALPYSSGTTGLPKGVMLTHKGLV 208
QY 265 RMAALVYGG---FRMRPNDIVDCLPLXH--SAGNIVGIGQCLLHGMTVWLKRFSAKR 318
DB 209 TSAQAQVDGDNPNLYMHSEDDVMICILPLEHIYSLNAVLCGG--LRAGVTILIMQRFDIYP 266
QY 319 FWDCKIKYNTIYOYIGELCRYLLNOPPREAENQHOVRMA-----LGNGLRQSIWTFNS 372
DB 267 FLELIQKYKVTIGFVPPPIVLAIAKSPVVDKYLSSVRTVMGSAAPLGRKELEDAV---R 322
QY 373 SRPHIQAFAEFGATECNCSLG-----NFDQVAGACGFSNRILLSFYPIRLVRVNE 424
DB 323 AKFPNAKLGQGYGTEAGPVVLAACLAFAKEPYEIKSGACG-----TVVRNAB- 369
QY 425 TMELIRPGDGVICPGPEGQVGLVRIQDPLRRFDGYLNOGANNKIAKDVFKKGDQA 484
DB 370 MKIVDPETNASLP--RNQGEI--CIRGQIMK--GYLNDPESTRTI-----DESG 415
QY 485 YL-TGVDVLYMDELGYLFRDRTGDTFRWKGENYSTTEVEGTLRLLDMADVAVYGEVPG 543
DB 416 WLHTGDIQDIDDDDELFIYDLRKEIKYKGFQVAPAELEALLTHPTISDAAV---VPM 471
QY 544 TEGRAG---NAAVASPTG--NCDLERAQVLEKELPLYARPIFLRLLPDLHKTGYKFKQT 599
DB 472 IDEKAGEVPVAVFVRINGFTTTEEEKIQFVSKQVYFKRIFRVFFVDAIPKSPSKILRK 531
QY 600 ELR 602
DB 532 DLR 534

Search completed: July 16, 2001, 18:15:13
Job time: 189 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: July 16, 2001, 18:15:11 ; Search time 18.37 Seconds
(without alignments)
1199.035 Million cell updates/sec

Title: US-09-405-504A-53
Perfect score: 3384
Sequence: 1 MLLGASLVGLVFLFSKLVLK.....RYVPLDQPAYSRIOAGEEKL 643

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2110	62.4	646	1	FATP_MOUSE
2	2087	61.7	646	1	FATP_RAT
3	1064	31.4	620	1	VLCS_HUMAN
4	1044	30.9	620	1	VLCS_RAT
5	1022	30.2	620	1	VLCS_MOUSE
6	714	21.1	623	1	FAT1_YEAST
7	350.5	10.4	522	1	CAIC_ECOLI
8	318	9.4	661	1	ACSA_COPCI
9	311.5	9.2	561	1	LCFA_ECOLI
10	303	9.0	545	1	4CL1_SOLTU
11	300	8.9	545	1	4CL1_SOLTU
12	299	8.8	547	1	4CL1_TOBAC
13	293.5	8.7	561	1	4CL3_ARATH
14	286	8.5	544	1	4CL1_PETCR
15	282	8.3	544	1	4CL2_PETCR
16	278.5	8.2	553	1	4CL1_VANPL
17	271.5	8.0	562	1	LCFA_HAFIN
18	271	8.0	543	1	FAT2_YEAST
19	270.5	8.0	656	1	ACSA_RHOCA
20	267.5	7.9	542	1	4CL2_TOBAC
21	262	7.7	694	1	ACSA_CRYPV
22	260	7.7	5255	1	BACA_BACLI
23	259	7.7	670	1	ACSA_EMENI
24	259	7.7	3587	1	SRLF_BACSU
25	255.5	7.6	3587	1	SRLF_BACSU
26	255	7.5	669	1	ACSA_PENCH
27	253	7.5	6359	1	BACC_BACLI
28	251	7.4	556	1	4CL2_ARATH
29	250	7.4	3587	1	TYCB_BACBR
30	249	7.4	548	1	YDID_ECOLI
31	247	7.3	675	1	ACSA_CANAL
32	246	7.3	713	1	ACSI_YEAST
33	246	7.3	2560	1	PPS2_BACSU

KW	Glycoprotein; Lipid transport; Transmembrane.
FT	TRANSMEM 14
FT	TRANSMEM 14
FT	TRANSMEM 115
FT	TRANSMEM 140
FT	TRANSMEM 169
FT	TRANSMEM 189
FT	TRANSMEM 293
FT	TRANSMEM 313
FT	TRANSMEM 537
FT	CARBOHYD 330
FT	CARBOHYD 330
FT	CARBOHYD 393
FT	CARBOHYD 518
SQ	SEQUENCE 646 AA; 71276 MW; 910B32BA8D985B

Query Match 62.4%; Score 2110; DB 1; Length 646;

[illegible]

QY 123 LASGDVAAFMENRNEFVGLWGLMAKLGVEAALINTNLRDALLHCLTTSRARALVFGSE 182

db 125 FAPGDWAVLEGRPEFVGLIGLAKAGVWAALINVINI BBEPLAFICITSAKVALIYCCE 184

189

183 MASAI CEVHASLDPSLSLFCSGSWEPAVPPSTEHLDP L L K DAP - KHL PSCPDKGFTDKL 241

db 185 MAAVAEVSEQLGKSLKFCGDLGPESILPDTQLLDPLAEAPTTPLAQAPGKGMDDL 244

242 EYLYTSGTTGIPKAAIVVHSRYRMAIVVVCERMPDNDIVVDCIPIVHSACNTVICGQ 301

[illegible]

245 FYIYTSGITGLPKAAIVVHSRYRIRIAAFGHHSYSMRAADVLYDCLPLYHSAGNIMGVQC 304

302 LLHGMTVIRKKFSASFWDCCIYKNCITVQYIGELCRYLLNQPPREAEHQVQRMALGN 361

db
305 VIYGLTVLRLKKFASRFWDCCVKYNTVVQYIGEICRYLLRQPVQDVEORHVRVLA VGN 364

[illegible]

36Z GLRQSIWTFSSRFFHIQVAEEFYGATECNCSLGNFDSQVACGFNSRILSFVPIRLVRV 421

365 GLRPAIWEFTQRFQVPQIGEFYGATECNCSIANMDGKVGSCGFNSRILTHVYPIRLVKV 424

422 NEDTMELIRGPDGVCIPCPDGPGEGLVGRIIOKDPIRRFDGYINOGANNKTIKDVKK 481

[illegible]

425 NEDTMEPLRDSEGLCPQPGEPGLLVGQINQQDPLRRFDGYVSDSATNKKIAHSVFRKG 484

482 DQAYLTGDVLVMDLGYLYFRDRTGDTFRWKGVNSTTEVEGTLRLLDMADVAVYGV 541

485 DSAYLSGDVLYMDELGYMYFRDRSGDTERWGENVSTTTEVAVLSRLIGOTDVAVYGAV 544

[illegible]

34Z PGTEGRAGMAAVASPTGNCNCLERFAQVLEKLELPYARPIFLRLPELHKGTGTYKFQKTEL 601

545 PGVEGKAGMAAIADPHSQDLPNSMYQELQVLSYARPIFLRLLPQVDTTGTFTKIQKTRL 604

602 RKEGFDPAIVKDPFLFYLDAAKGRVPLDOEAYSRIQAGEEKL 643

C10 PUNTO D'AVANGUARDIA CON IL SUO STILE DI VITA

505

96

605 QREGFDPRTSDRLFFLDLKGRYVPLDERVHARICAGDFSL 646

3
5

FATP_RAT

ID	FATP_RAT	STANDARD;	PRT; 646 AA.
1	007010.		

AC P97849;
DTT 01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)

30-MAY-2000 (REL. 39, Last annotation update)
LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).

SLC27A1 OR FATP.

OS *Rattus norvegicus* (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

11/11/2023 10:11:11 AM

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98041635; PubMed=9375787;
 RA Schaap F.G., Hamers L., van der Vusse G.J., Glatz J.F.C.;
 RT "Molecular cloning of fatty acid-transport protein cDNA from rat.";
 RL Blochim Biophys. Acta 1354:29-34(1997).
 CC -1- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS
 CC ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
 CC AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
 CC IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
 CC TRIGLYCERIDE SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U89529; AAC53424.1; -
 CC InterPro: IPR000873; -
 CC Pfam: PF00501; AMP-binding; 1.
 CC PROSITE: PS00455; AMP-BINDING; 1.
 CC Glycoprotein; Lipid transport; Transmembrane.
 CC FT TRANSMEM 14 34 POTENTIAL.
 CC FT TRANSMEM 115 135 POTENTIAL.
 CC FT TRANSMEM 140 160 POTENTIAL.
 CC FT TRANSMEM 169 189 POTENTIAL.
 CC FT TRANSMEM 293 313 POTENTIAL.
 CC FT TRANSMEM 537 557 POTENTIAL.
 CC FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 518 518 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 646 AA; 71283 MW; C450CF174CC2EB87 CRC64;
 DR
 DR
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding; 1.
 DR PROSITE: PS00455; AMP-BINDING; 1.
 DR Glycoprotein; Lipid transport; Transmembrane.
 DR FT TRANSMEM 14 34 POTENTIAL.
 DR FT TRANSMEM 115 135 POTENTIAL.
 DR FT TRANSMEM 140 160 POTENTIAL.
 DR FT TRANSMEM 169 189 POTENTIAL.
 DR FT TRANSMEM 293 313 POTENTIAL.
 DR FT TRANSMEM 537 557 POTENTIAL.
 DR FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
 DR FT CARBOHYD 393 393 N-LINKED (GLCNAC. .) (POTENTIAL).
 DR FT CARBOHYD 518 518 N-LINKED (GLCNAC. .) (POTENTIAL).
 DR SEQUENCE 646 AA; 71283 MW; C450CF174CC2EB87 CRC64;
 SQ
 Query Match 61.7%; Score 2087; DB 1; Length 646;
 Best Local Similarity 60.4%; Pred. No. 7.3e-166;
 Matches 388; Conservative 98; Mismatches 154; Indels 2; Gaps 2;
 4 GASLVGVLLFSKL-VLKLPTQVQFSLFLYLGGSGWRFRIRVFIKIRIDIFGLVLLKV 62
 5 GAGTASVASLGLLWLLGLPWTSAARAFVYVGGSGWRFLRIVCKTARRDLFGLSVLRV 64
 63 KAKVROCLQERRVPIILFASTVRHDPDKTALIFEGTDTHTFRQLDEYSSVANFLQARG 122
 65 KLEURRHRAGDTIPRIFAQVQRPRLALVDASSGICWTFAQLDTYISVANANFLQLG 124
 123 LASGDVAAIFWENRNEFVGLWGLMAKLGVEAALINLRDALLHLGTLTSRARALVFGSE 182
 125 FAPGDVAVFLEGRPEFVGLWGLARAGVVAALLNLRREPLAFCLGTSAAKALIYGGE 184
 183 MASAICEVHASLDPSSLFCSSGWEFGAVPPSTEHLDPDLKADP-KHLPSCDPKGTGDKL 241
 185 MAAVAEVSQGLKSLKFCGSDIGLPSVLPDQLDPLMLAEAPTTPLAQAPGGMDDRL 244
 242 FYIYTSGETTGLPKAAIVVHSRYRMAALVYVGFMRPNDIVDCLPLYHSAGNIVGIGOC 301
 245 FYIYTSGETTGLPKAAIVVHSRYRMAALVYVGFMRPNDIVDCLPLYHSAGNIVGIGOC 304
 302 LLHGMTVWIRKKSASRFDWDCIKYNTIVQYIGELCYRLNPPREAEHNOHVRMALGN 361
 305 IYGLTVVLRKKSASRFDWDCIKYNTIVQYIGELCYRLNPPREAEHNOHVRMALGN 364
 362 GLROSINWTFSSRPHIPQVAEYFATGECNCSLGNFOSVQACGFCNRSILSVFYPIRLRV 421
 365 GLPAIWEETQGVGRVQIGEFYATGECNCSIANMDGKVGSCGFCNRSILSVFYPIRLRV 424

QY 422 NEDTMELIRPGDYVICPCQPGEGQLVGRILQKDLRRFDGYLNQGNKKIAKDVFKKG 481
 DB 425 NEDTMELIRDSQGLCIPQPGEGGLVGRILQKDLRRFDGYVSDSNKKNIAHSVFKKG 484
 QY 482 DQALVTGDLVMDLGYLYFRDRTGDTFRWKGENVSTVEGTLRLSLMDADVAVIGVEY 541
 DB 485 DSAVLSGDLVMDLGYLYFRDRTGDTFRWKGENVSTVEGTLRLSLMDADVAVIGVEY 544
 QY 542 PGTGRAGMAVASPTGNCDLERFAQVLEKELPIYARPIFLRLPELHKYTGTFKQTEL 601
 DB 545 PVEGKSGMAIAADPNLQNDPNSMYQELQVLAQYAFIPLRLPQVDTTGTGFKIQTRL 604
 QY 602 RKEGFDPAIVKDFLYLDAOKGRVYVPLDOEAYSRIQAGEEKL 643
 DB 605 QREGFDPRQTSRDLFFLDLQKGRYLPIDRVERHARICAGDFSL 646

RESULT 3

VLCS_HUMAN STANDARD; PRT; 620 AA.
 AC 014975;1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
 DE FATTY-ACID-COA LIGASE).
 GN FACVLI OR VLCS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Uchiyama A., Aoyama T., Kamiyo K., Wakui K., Fukushima Y.,
 RA Shimozawa N., Suzuki Y., Kondo T., Orii T., Hashimoto T.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95216327; PubMed=10198260;
 RA Steinberg S.J., Wang S.J., Kim D.G., Mihalik S.J., Watkins P.A.;
 RT "Human very-long-chain acyl-CoA synthetase: cloning, topography, and
 RT relevance to branched-chain fatty acid metabolism.";
 RL Biochem. Biophys. Res. Commun. 257:615-621(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, PLACENTA AND
 CC PANCREAS.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D88308; BAA23644.1; -
 CC EMBL: AF096290; AAC64973.1; -
 CC MIM: 603247; -
 CC InterPro: IPR000873; -
 CC Pfam: PF00501; AMP-binding; 1.
 CC PROSITE: PS00455; AMP-BINDING; 1.
 CC Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
 CC FT TRANSMEM 1 21 POTENTIAL.
 CC FT TRANSMEM 107 127 POTENTIAL.
 CC FT TRANSMEM 262 282 POTENTIAL.
 CC SEQUENCE 620 AA; 70312 MW; BFD3E1BB67B3DF8 CRC64;
 DR
 DR EMBL: D88308; BAA23644.1; -
 DR EMBL: AF096290; AAC64973.1; -
 DR MIM: 603247; -
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding; 1.
 DR PROSITE: PS00455; AMP-BINDING; 1.
 DR Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
 DR FT TRANSMEM 1 21 POTENTIAL.
 DR FT TRANSMEM 107 127 POTENTIAL.
 DR FT TRANSMEM 262 282 POTENTIAL.
 DR SEQUENCE 620 AA; 70312 MW; BFD3E1BB67B3DF8 CRC64;
 SQ

Query Match

31.4%; Score 1064; DB 1; Length 620;

AC Q35488;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
 DE FATTY-ACID-COA LIGASE).
 GN FACVL1 OR VLACS OR VLCS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Kemp S., Lu J.-F., Smith K.D.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; AF033031; AAB87982.1; -
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 SQ SEQUENCE 620 AA; 70366 MW; 77C98BD0DE3B9FFB CRC64;

Query Match 30.2%; Score 1022; DB 1; Length 620;
 Best Local Similarity 37.7%; Pred. No. 3.2e-77;
 Matches 235; Conservative 108; Mismatches 248; Indels 32; Gaps 12;

QY 32 LYLGGGWRIRVFIRK-----IRDFGLGLVLLKVKAKVQCLOER--RTVPILFASTVR 85
 DB 5 LYTGLALLPLLLTCCCYLLQDVRYFLQLANARRVRSYRQRPVRLIRAFLEQAR 64
 QY 86 RHPDKTALIFEQDTHTWTFRLQDEYSSSVANFLQAR--GLASGDVAATFMENRNEFYGLWL 144
 DB 65 KTHPKPELLER--DETLYAQVDRRSNOVARALHDQLGLRQGDVAFMGNEPAYVWIWL 122
 QY 145 GMKGLGVEALINLNRDALLCLTTSRARALVFGSEMASACEVHASL--DPSLSLFC 202
 DB 123 GLLKLGCPMACLNYNIRAKSLHFCQCGAKVLLASPDLOEAEEALPTLKDDAVSVFYV 182
 QY 203 SGSEPCAVPSPVEHEDPL-LKDAKHLPCSPDKGTFKLFYIYTSGTGLPKAAIIVVHS 261
 DB 183 SRTSNTNGVDTLIDKVDGVAETPESWRS--EVTETTPAVIYTSGTGLPKAAIINH 240
 QY 262 RYRMAALVYVYGRM-----RPNDIVYDCLPLYSAGNTVIGTQCQLLHGTMVIRKFS 315
 DB 241 R-----LWYGTGLAMSSGITAQDVYITTMPLHYSAALMIGLHGCIVVGATLALRSKS 293
 QY 316 ASRFWDICIKYNTIVYICELCYLLNLOPPREAHNOHVRMALGNLQSGIWTNFSRRF 375
 DB 294 ASQFWDICRYNTIVYICELCYLLNLOPPREAHNOHVRMALGNLQSGIWTNFSRRF 353
 QY 376 HIPOVAEFGATECNCSLGNFDSQVACGFSNRILSFVYPIRLVRYNEDTMELIRGPDGV 435
 DB 354 GDHIVYEFYASTEIGNICFVNPYPRKIGAVGRANVYLRQKRYELIKYDVKEDEPVRDANGY 413
 QY 436 CIPQCPPEPQGLVGRITIQKPLRFDFGYL--NQANNNKKTAKDVFKKGDDQAYLTGDLVMD 494

DB 414 CIKVPKGEVGLLVCKITQLTP---FIGVAGGKTOTEKKKLRDVKKGIYFNSGDLLMID 470
 QY 495 ELGILYFRDRTGDFTRWKGENVSTTEYEGTLRLMDADVAVYGEVPGTEGRAGMAAV- 553
 DB 471 RENFVYFDHVRGDFTRWKGENVSTTEYEGTLRLMDADVAVYGEVPGTEGRAGMAAV- 530
 QY 554 ASPTGNCOLERPAQVLEKEKELPLARPIFLRLPLDELHKTGYKFKOKTELRKEGPDPAIVKD 613
 DB 531 IKENYEFNGKGLFOHIAEYLPSPARFLRQDTIETITGTFKKRKVTLMEEGFNPTVIKD 590
 QY 614 PLFYLDACKGRYVPLDQEAYSRI 636
 DB 591 TLYFMDDAEKTFVPMENIYNAI 613

RESULT 6
 FATL_YEAST STANDARD; PRT; 623 AA.
 AC P38225;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.
 GN FATL OR YBR041W OR YBR0411.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
 CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Vissers S.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION.
 RC STRAIN=W303A;
 RC MEDLINE-97236810; PubMed-9079682;
 RA Faergeman N.J., Dirusso C.C., Eiberger A., Knudsen J., Black P.N.;
 RT "Disruption of the Saccharomyces cerevisiae homologue to the murine
 RT fatty acid transport protein impairs uptake and growth on long-chain
 RT fatty acids.";
 RL J. Biol. Chem. 272:8531-8538(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE. AND
 CC THIS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY
 CC PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN
 CC UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY
 CC ACID SYNTHESIS IS COMPROMISED.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
 CC BINDING OF AMP TO THEIR SUBSTRATE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; Z35910; CA84983.1; -
 DR PIR; S45899; S45899.
 DR SGD; S0000245; FATL.
 DR InterPro: IPR000873; -
 DR Pfam; PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Lipid transport; Transmembrane.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 54 71 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC...) (POTENTIAL).

QY 109 EYSSVANFLOARGLASGDVAIAEMENREFFVGLWMAKLGVEAALINTNLRDALLHC 168
Db 50 QETNRTANFYTLGIRKGDVHALHDCPEFIFCFGLAKIGAIMVPIINARLLCEESAWI 109
QY 169 LTTSRARALVFGSEMASACEVHASLDPSLSLFCSGSWEPGAVP--PSTEHLDPDLLKQAP 226
Db 110 LQNSQACLLVTSQAQFPMYQQIQEEDATQRLHC---LTDVALPADDGVSSFTQLKNQOP 166
QY 227 KHLPCSPDKGFTDKLFYIYTSGTGLPKAAIVVHSRYRYMAALVYVG---FMRPNDIVY 283
Db 167 ATLCYAPPLSTDDTAEILFTSGTTSRPGKWITH---YNLREAGYYSAWQCALRDDVYL 223
QY 284 DCLPLVHSAGNIVGICQLLHGMVTVIRKFSASRFWDCIKYNTIVQYIGELCRYLLN 343
Db 224 TVMPAFHIDQCQTAAMAFSAGATFVLVEKYSARAFWQGVQRYRATVTECIPMTIRLMV 283
QY 344 OPPEAENHOVRMALGN--GLRQSIWTFNSRFHPOVAEFGYATGECNCS--LGNFDS--- 398
Db 284 QPSANDQOHLREVMFYLNLSQEKDAFCERFV--RLTSGMVTETIVGIIIGDRPGDKR 342
QY 399 ---OVGACGFSRLSFVYPIRLVRVNDTMDLIRGPDG--VCIPCQPGEPGLVRIQK 454
Db 343 RWPISGRVGF-----CYEAER---DDHNRPLPAGEIGEICI---KGIPGKTIK--- 386
QY 455 DPLRRFDGVLNQCANNKKTAKDVFKGDAQYL--TGDVLYMDELGYLYPRDRTGDTFRWKG 513
Db 387 -----EYFNQPTAKVLEAD-----GWLHTGDTGVRDEEDFFYVDRRCNMIKRGG 433
QY 514 ENVSTTEVECTLSRLDMDADVAVGVVEVPGTEGRAGMAAASVPTGNCDLERFAQVLEKEL 573
Db 434 ENVSCVELENIAAHPKIQDVIIVVGIKDSIRDEAIKAFVVLNEGETLSSEEFRCQNM 493
QY 574 PLYARPIFLRLPELHKHTGTGYKFTKTEL 602
Db 494 AKFKVPSYLEIRKDLPRNGSGKIIRKNLK 522

RESULT 8
ACSA_COPCI STANDARD; PRT; 661 AA.
AC 013440;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1) (ACETATE--COA LIGASE) (ACYL-
DE ACTIVATING ENZYME).
GN ACS-1.
OS Coprinus cinereus (Inky cap fungus).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OC Coprinaceae; Coprinus.
OX NCBI_TaxID=5346;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JV6;
RA Chaire P.T., Casselton L.A., Connerton I.F.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: ATP + ACETATE + COA -> AMP + DIPHOSPHATE +
CC ACETYL-COA.
CC -I- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y15417; CAA75612.1; -.
DR InterPro; IPR000873; -.
DR Pfam; PF00501; AMP-binding; 1.

DR PRINTS: PR00154; AMP-BINDING.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Ligase.
SQ SEQUENCE 661 AA; 73094 MW; D647C974095795D9 CRC64;

Query Match 9.4%; Score 318; DB 1; Length 661;

Best Local Similarity 25.3%; Pred. No. 1.5e-18;
Matches 146; Conservative 90; Mismatches 240; Indels 100; Gaps 25;

QY 86 RHPDKALIFE---GTDTHWTFQLDDEYSSVANFLOARGLASGDVAIAEMENREFFV 141
Db 96 KHPEKTAIIEADPEGEGREISYAEALLREVCSIANVLKSGVKKGDVSVVLPMTQAVA 155
QY 142 LNLGMALGVEAALINTNLRDALLHCLTTSRARALVFGSE-----MASAICE 189
Db 156 AFLACARIGAIHSVVFAGFAEALRDMQCKSRVLTSDEGRGKAIATKAIDAALK 215
QY 190 VHASLDPSLSLFCSGSWEPGAVPSTEHLDPLLDKAPKHLPS--CPDK--GFTDKLFYIT 246
Db 216 ECPAVEKVLVLRKTN-----PVPWTEGRDWWHEAVRVPYCPPEVMASEDPLFLYIT 270
QY 247 SGTTLGPKAAIVVHSRYRYMAAL--VYGFMRPNDIVDCLPLYHSAGNIVG-----IGQ 300
Db 271 SGSTGKPGVVHTTGGYLLCAALTIVYVDFVHDD--RFACMA---DVGWITGHTYIVGP 326
QY 301 CLLHGMVWIRKK---PSASRFWDCIKYNTIVQYIGELCRYLLNOPPREAENH--- 353
Db 327 LAIGATTVFESTPVYPTPSRYMETVEKYLQTP--YSAPTALRLRLGHEHVNKHDLSS 385
QY 354 -QVRMALGNLGRSIWTFNSRFHPOVAEFGYATGECNCSLGNFDSQVAGCGFNSRLSF 412
Db 386 LRVLSGVGEPIINPAWHYN-----EHVKTECAIVDTFWQTETGS-----IVVT 430
QY 413 VYPIRLVRVNDTMDLIRGPDGVCIPCQPGEPGL-----VGRIOKDLPRFDGYLNQAN 469
Db 431 PFGAI-----ETK-----PGAATVPFFGIEPAILETTGKVLGENDV---EGLVITAH 477
QY 470 NKKTAKDVFKGD--QAYL-----TGDVLYMDELGYLYPRDRTGDTFRWKG 516
Db 478 WPSARTIY---GDHQRYLEYMKPYGTYTGGGAARDEGYIWKGRVDVIVNSGRL 535
QY 517 STTEVECTLSRLDMDADVAVGVVEVPGTEGRAGMAA--SPTGNCDLERFAQVLEKEL 575
Db 536 STAEISALITHTGVAETAVIGTADLT--GQAVYAVFTLKPPEKFEADENAG--LSKELIL 593
QY 576 -----YARPIFLRLPELHKHTGTGYKFTKTELK 603
Db 594 QVRKIIGPFAAPKRIYIVSDLPKTRSGKIMRILRK 629

RESULT 9
LCFA_ECOLI STANDARD; PRT; 561 AA.
ID LCFA_ECOLI
AC P29212;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2.1.3) (LONG-CHAIN ACYL-COA
DE SYNTHETASE).
GN FADD OR OLD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=94150456; PubMed=8107670;
RA Fulda M., Heinz E., Wolter F.P.;
RT "The fadD gene of Escherichia coli K12 is located close to rnd at
RT 39.6 min of the chromosomal map and is a new member of the
RT AMP-binding protein family.";

RESULT 12
4CL1_TOBAC

ID 4CL1_TOBAC STANDARD; PRT; 547 AA.

AC 024145; PRT; 547 AA.

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE 4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA SYNTHASE 1).

GN 4CL1

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxID=4097;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96416441; PubMed=88193324;

RA Lee D., Douglas C.J.;

RT "Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins.";

RL Plant Physiol. 112:193-205(1996).

CC -|- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE + 4-COUMAROYL-COA.

CC -|- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.

CC -|- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; U50845; AAB18637.1; .

CC DR HSP; P08659; 1LC1.

CC DR Mendel; 24477; Nicta; 1179; 24477.

CC DR InterPro; IPR000873; .

CC DR Pfam; PF00501; AMP-binding; 1.

CC DR PROSITE; PS00455; AMP_BINDING; 1.

CC DR Ligase; Phenylpropanoid metabolism; Multigene family.

CC SK SEQUENCE 547 AA; 59842 MW; 154DF6D684E3F51B CRC64;

CC

Query Match 8.8%; Score 299; DB 1; Length 547;

Best Local Similarity 24.0%; Pred. No. 4.2e-17;

Matches 132; Conservative 93; Mismatches 232; Indels 92; Gaps 22;

QY 90 KTALEFGDTHTFRLQDEYSSVAFNLQARGLAGSDVAALFENRNEFFVGLWGLMAKL 149

DB 46 RCLINGANDQIYTAELVETLCRKVAVGLNKLGIQOKDTIMLLPNSPEFFVFAEMGASYL 105

QY 150 GVEAALINTLRDALLHCLTTSRALVFGSEMASAICEVHASLDPSLSLFCSGSEWPG 209

DB 106 GAISTMANPLFPAEWVKQAKASSAKIITQSCFVGKVDYASEND--VKVICIDSAPEG 163

QY 210 A-----VPPSTHLDPLKAPKHLFPCSPDGFDTKLYIYTSGTGLPKAAIVVHSRY 264

DB 164 CLHFSLTQSDHEIPEVKIQDDVVALP-----YSSGTGLPKGVMLTHGLV 212

QY 265 RMAALVYXG-----FRMRPNIDVDCPLPHSHSAGNIVGICQLLHGM-----TWVIRKFFSA 316

DB 213 TSVAQVDGENANLYMHSEDLVLCVPLFH---IYLSNLSILLGLRVGAAILIMQKFDI 268

QY 317 SRFWDCLIKYNTVOYIGELCRYLLNQPPEAENQHVRMA-----LGNGLRQSIWTN 370

DB 269 APFLELIQKYKVSFPFPVPLAIAKSPIDVSDYLSVTVMSGAAPLGKELEDAVTK 328

QY 371 FSSRRHIFQVAFFYGATCNCSLG-----NFDQVGACGCFNSRILSFVPIRLVRVN 422

Db 329 FPN-----AKLGQYGMTEAGPVLAMCLAFAPEDIKSGACG-----TVVRNA 372

QY 423 EDTMELIRGPGVCIPCPGPGGOLVGRIOKDLRRDGYLNO--GANNKKIADKVPKG 481

Db 373 E--MKIVPDPDTCGLP--RNQFGEI--CIRGQDMK--GYLNDPEATRTIDK-----417

QY 482 DQAYL--TGDVLVMDLELYLFRDTRFRWKGENVSTVEGTLSRLLDMAVAVYGV 540

Db 418 -EGWLHTGDIGFIDEDELFTVDRKLKELIKYKGFQVAPAEATEALLNHPNISDAV-----472

QY 541 VPGTEGRAG---MAAVASPTGNCOLE--RFAQVLEKELPLY---ARPLRLLLPELHGTGT 593

Db 473 VPMKDEQAGEVFPVAFVVRNSGSAITEDVDFISKVIFYKRVKRVFVETVP---KSPS 529

QY 594 YKFORTELR 602

Db 530 GKILRKDLR 538

RESULT 13

4CL3_ARATH

ID 4CL3_ARATH STANDARD; PRT; 561 AA.

AC 095777;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE 4-COUMARATE--COA LIGASE 3 (EC 6.2.1.12) (4CL 3) (4-COUMAROYL-COA SYNTHASE 3).

GN 4CL3 OR AT1G65060 OR F16G16.6.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=10417722;

RX Ehling J., Buettner D., Wang Q., Douglas C.J., Somssich I.E., Kombrink E.;

RA "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms.";

RT Plant J. 19:9-20(1999).

RL [2]

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Ruehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A., Millschler J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shing P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Xu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";

RL Nature 408:816-820(2000).

CC -|- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE + 4-COUMAROYL-COA.

CC -|- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.

CC -|- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC

7